UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/601,258	06/19/2003	Wolfgang Stephan	34874-165 UTIL	6725

64280 7590 03/31/2008

MINTZ, LEVIN, COHN, FERRIS, GLOVSKY & POPEO, P.C.
ATTN: PATENT INTAKE CUSTOMER NO. 64280
ONE FINANCIAL CENTER
BOSTON, MA 02111

EXAMINER	
COLAN, GIOVANNA B	
ART UNIT	PAPER NUMBER
2162	
MAIL DATE	DELIVERY MODE
03/31/2008	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/601,258	STEPHAN, WOLFGANG	
	Examiner	Art Unit	
	GIOVANNA COLAN	2162	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *31 October 2007*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-6,8-42 and 44-51* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-6,8-42 and 44-51* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date *06/19/2003*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. This action is issued in response to the Amendment filed on 10/31/2007.

2. Claims 1, 15, 25, 37, and 51 were amended. Claims 7, and 43 were canceled. No claims were added.

3. Claims 1 –6, 8 – 42, and 44 – 51 are pending in this application.

4. Applicant's request for reconsideration of the finality of the rejection of the last Office action dated 09/28/2007 is persuasive and, therefore, the finality of that action is withdrawn.

Response to Arguments

5. Applicant's arguments with respect to claims 1 –6, 8 – 42, and 44 – 51 have been considered but are moot in view of the new ground(s) of rejection.

Information Disclosure Statement

6. The information disclosure statement (IDS) was submitted on 06/19/2003. The submission is in compliance with the provisions of 37 CFR 1.97. Accordingly, the information disclosure statement is being considered by the examiner.

Claim Rejections - 35 USC § 101

7. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

8. Claims 25 – 60 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter.

Claim 25 (which recites; "An inverted index for a collection of documents....") is directed to non-statutory subject matter, since it claims functional descriptive material, per se (data structure, per se).

Both types of “descriptive material” are nonstatutory when claimed as descriptive material per se, 33 F.3d at 1360, 31 USPQ2d at 1759. When functional descriptive material is recorded on some computer-readable medium, it becomes structurally and functionally interrelated to the medium and will be statutory in most cases since use of technology permits the function of the descriptive material to be realized. Compare In re Lowry, 32 F.3d 1579, 1583-84, 32 USPQ2d 1031, 1035 (Fed. Cir. 1994)(discussing patentable weight of data structure limitations in the context of a statutory claim to a data structure stored on a computer readable medium that increases computer efficiency) and >In re< Warmerdam, 33 F.3d *>1354,< 1360-61, 31 USPQ2d *>1754,< 1759 (claim to computer having a specific data structure stored in memory held statutory product-by-process claim) with Warmerdam, 33 F.3d at 1361, 31 USPQ2d at 1760 (claim to a data structure per se held nonstatutory).

Claims 37, and 51 recite “An article comprising a machine readable medium storing instructions”, which appears to be directed to non-statutory subject matter ([0062], specification, “…any **computer product**, apparatus and/**or** device...including machine-readable medium...as a **machine-readable signal**…”).

Claim Rejections - 35 USC § 103

9. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

10. Claims 1 – 6, 9 –14, 16 – 30, 32 – 42, and 45 – 50, and 52 – 60 are rejected under 35 U.S.C. 103(a) as being unpatentable over Broder et al. (Broder hereinafter) (US Patent Application Pub No. 2004/0243560 A1, filed: May 30, 2003) in view of Moffat et al. (Moffat hereinafter) (Non-Patent Literature: "Self-Indexing Inverted Files for Fast Text Retrieval"; Alistar Moffat, and Justin Zobel; February 1994, 1994 Australian Database Conference, and 1994 IEEE Conference on Data Engineering).

Regarding Claim 1 and 37, Broder discloses an article comprising a machine-readable medium storing instructions operable to cause one or more machines to perform operations comprising:

determining a first value x such that x is determined as an integer corresponding to a quantity of documents including at least a majority of the index terms (Page 17 and 18, [0307] and [0314], lines 1 – 3 and 3 – 9; respectively, Broder[1]).

[1] Wherein the top "n" documents corresponds to the documents with the majority of index terms claimed; specifically "n" corresponds to the value x claimed. The scoring procedure utilized by Broder (as disclosed in detailed in Page 16, [0286], lines 5 – 6, Broder) including terms associations with upper bounds on its

determining a second value y, where y does not exceed x, (Page 15, [0277], lines 1 – 2, **k be the smallest** index, Broder).

However, Broder does not explicitly disclose that: x is representative of a first location **for inserting** a first skip entry in an inverted index; and that the second value y determined as an integer and representative of a second location **for inserting** a second skip entry in the inverted index. On the other hand, Moffat discloses that: x is representative of a first location **for inserting** a first skip entry in an inverted index (Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – **are introduced into the compressed inverted list.** For example, suppose that p1 synchronisation points are allowed. Then the index into the inverted list contains p1 'document number, bit address' pairs, and can itself be stored as a compressed sequence of 'difference in document number, difference in bit address'…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat; also see Page 18, Section 4.4 Implementation, 1st paragraph, "…**inverted files were built with skips inserted** into each list assuming that k had some fixed value …", Moffat); and that: the second value y determined as an integer and representative of a second location **for inserting** a second skip entry in the inverted index (Page 14, 1st and 2nd paragraph, "…the inverted list becomes a sequence of blocks of three pairs each, with skips separating the

---

maximal contribution to documents scores; wherein the maximal contributing of index terms corresponds

blocks…", and "…a2 is the address of the first bit of the second skip pair, a3 is the address of the first bit of the third skip, and so on…", Moffat; and also see Page 14, 4th paragraph, "a Golomb code is used for both the **inverted lists and the two components of the skips inserted into each**.", Moffat). It would have been obvious to one of ordinary skill in the art at the time the invention was made to incorporate the Moffat's teachings to the system Broder. Skilled artisan would have been motivated to do so, as suggested by Moffat (Page 15 and 16, paragraph 5, and 2; respectively, Moffat), to insert skips and compress inverted lists to allow both disk space and query processing time to be reduced by having a faster inverted file index by fetching fewer terms and list. In addition, both of the references (Broder and Moffat) teach features that are directed to analogous art and they are directed to the same field of endeavor, such as, databases management systems, inverted lists, skipping, and indexing. This close relation between both of the references highly suggests an expectation of success.

Furthermore, the combination of Broder in view of Moffat discloses:

generating the inverted index for the collection of documents (Page 9, [0162], lines 1 – 3, Broder; and Page 1, Abstract, "…an internal index in each inverted list…", Moffat), the inverted index including an inverted list for each of the index terms (Page 14, [0244], lines 2 – 3, Broder), each inverted list including at least one posting (Page 14, [0244], lines 2 – 5, Broder) and, if the number of postings exceeds x (Page 14,

---

to the majority of the index terms as claimed. Additionally, the top n scoring documents corresponds to a quantity of documents as claimed.

[0245], lines 7 – 10, Broder[2]), further including the first skip entry **inserted** after the xth posting (Page 14, [0245], lines 10 – 12, Broder; Page 13, Section: 4.1 Skipping, "…When k<<p, faster performance is possible if synchronization points-additional locations at which decoding can commence-are introduced into the compressed inverted list…", Moffat; and also see Page 18, Section 4.4 Implementation, 1st paragraph, "…**inverted files were built with skips inserted** into each list assuming that k had some fixed value …", Moffat) and one or more second skip entries **inserted** thereafter at intervals of every yth posting (Page 15, [0277], lines 1 – 6, Broder[3]; Page 14, 1st and 2nd paragraph, "…the inverted list becomes a sequence of blocks of three pairs each, with skips separating the blocks…", and "…a2 is the address of the first bit of the second skip pair, a3 is the address of the first bit of the third skip, and so on…", Moffat and also see Page 14, 4th paragraph, "a Golomb code is used for both the **inverted lists and the two components of the skips inserted into each**.", Moffat);

wherein:

the at least one posting includes a document identifier identifying a document in the collection of documents (Page 14, [0244], lines 1 – 6, document unique identifier DID, Broder; and Page 3, Section 2: Document Database, 1st paragraph, "Each document is known by a unique indetifier…", Moffat);

---

[2]Broder discloses that the method next(id) (page 14, [0245], lines 8 – 10). This method states that if there is **not** such document which DID >=id (where DID = number of documents, and id = number of posting), then the term iterator returns a special posting that is larger than all the existing DIDs. This implies that, if there is DID < id (the number of postings exceeds the number of documents entered), then the iterator returns the special posting mentioned above.

[3] The cursor is advanced to the position of k value. There is a skip at the kth value.

the first and second skip entries including a document identifier that is included in a boundary posting of a block of postings immediately adjacent to the skip entry in the inverted list (Page 15, [0277], lines 1 – 6, Border[4]; and Page 13, Section 4.1 Skipping, 6th paragraph, "… For example, suppose that p1 synchronisation points are allowed. Then the index into the inverted list contains p1 'document number, bit address' pairs, and can itself be stored as a compressed sequence of 'difference in document number, difference in bit address'…"; Moffat), where a block of postings includes postings having document identifiers ranging from a lower to an upper value and where a boundary posting is a posting having a document identifier of either the lower or the upper value (Page 16 and 18, [0286] and [0314], lines 5 – 9 and 3 – 9; respectively, Broder).

The combination of Broder in view of Moffat also discloses:

wherein y does not equal x, such that the first and second skip entries are provided at different intervals (Fig. 13, see jump/skip entry after "software" from 5-12, and jump/skip entry after "is" from 14 - 15; as shown in Fig 13, the intervals are different; Broder; and also Page 20, lines 9 – 15; "Another way to include skipping in an index is to vary the L parameter for each inverted list rather than use the same value…"; Moffat).

[4] Wherein doc_i corresponds to the document identifier claimed. In addition, the step of advancing the cursor to next doc_I k corresponds to the step of

Regarding Claims 2, and 38, the combination of Broder in view of Moffat discloses a method, wherein each posting further includes a position identifier identifying a position of the index term in the document (Page 14, [0244], lines 8 – 9, offsets of occurrences, Broder), and a frequency of the index term occurring in the document (Page 14 and 18, [0244] and [0316], lines 7 – 8 and 9 – 11, number of occurrences of the terms/ frequency; respectively, Broder).

Regarding Claims 3, and 39, the combination of Broder in view of Moffat discloses a method, wherein the boundary posting includes a document identifier having the lower value in the range of values and the block of postings follow the first skip entry in the inverted list (Fig. 27, item 5, Page 16 and 17, [0301], lines 10 – 18, Broder[5]; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

Regarding Claims 4, and 40, the combination of Broder in view of Moffat discloses a method, wherein the first skip entry further includes information to locate the

[5] According to Broder, the pivot DID is the smallest DID that might be a candidate (Page 17, [0301], lines 17 – 18). In addition, Broder discloses that this pivot term goes through an "if" statement which finds a first pivot term with UB (upper bound) greater than the threshold (Fig. 27, item 5). This implies that the "next" method will return the smallest possible document number following the last one. In addition, the smallest document number corresponds to the lower value claimed.

next skip entry in the inverted list (Fig. 27, items 13 and 22, Broder[6]; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

Regarding Claims 5, and 41, the combination of Broder in view of Moffat discloses a method, wherein the boundary posting includes a document identifier having the higher value in the range of values and the block of postings precede the first skip entry in the inverted list (Page 17, [0302], lines 22 – 26, Broder[7]; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

Regarding Claims 6, and 42, the combination of Broder in view of Moffat discloses a method, wherein the first skip entry further includes information to locate the

[6] Posting[aterm] (located in the function next()) represents the information to locate next skip.
[7] The next() function iterates trough the list and selects from the preceding terms the term with the location greater (largest document number of documents) than the pivot location. Wherein the location greater (largest document number of documents) than the pivot location corresponds to the higher value as claimed.

next skip entry in the inverted list (Fig. 27, items 13 and 22, Broder[8]; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

Regarding Claims 9, and 45, the combination of Broder in view of Moffat discloses a method, wherein the collection of one or more documents includes one or more binary files, data tables, source code files, text documents or combinations thereof (Page 9, [0158], lines 1 – 13, Broder).

Regarding Claims 10, and 46, the combination of Broder in view of Moffat discloses a method including all the limitations of claim 1, and 37, as disclosed above, further comprising:

compressing the inverted index (Page 15, [0273], lines 1 – 3, zipping, Broder).

Regarding Claims 11, and 47, the combination of Broder in view of Moffat discloses a method, wherein substantially all of the index terms occur in x documents or fewer (Page 15, [0257], lines 14 – 16, Broder).

[8] Posting[aterm] (located in the function next()) represents the information to locate next skip.

Regarding Claims 12, and 48, the combination of Broder in view of Moffat discloses a method, wherein at least approximately 80% of the index terms occur in x documents (Page 17, [0307], lines 1 – 3, top n results, Broder).

Regarding Claims 13, and 49, the combination of Broder in view of Moffat discloses a method, wherein for each inverted list, if the number of postings exceeds x, further including a skip entry before the first posting in the inverted list (Page 15, [0257], lines 12 – 20, the result is inserted, Broder; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

Regarding Claims 14, and 50, the combination of Broder in view of Moffat discloses a method, wherein for each inverted list, if the number of postings exceeds x (Page 14, [0245], lines 7 – 10, Broder[9]), further including a skip entry after the last posting in the inverted list (Page 14, [0245], lines 10 – 12, Broder; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first

[9]Broder discloses that the method next(id) (page 14, [0245], lines 8 – 10). This method states that if there is **not** such document which DID >=id (where DID = number of documents, and id = number of posting), then the term iterator returns a special posting that is larger than all the existing DIDs. This implies that, if

skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

Regarding Claims 16, and 52, the combination of Broder in view of Moffat discloses a method, wherein each posting further includes, a position identifier identifying a position of the index term in the document (Page 14, [0244], lines 8 – 9, offsets of occurrences, Broder), and a frequency of the index term occurring in the document (Page 14 and 18, [0244] and [0316], lines 7 – 8 and 9 – 11, number of occurrences of the terms/ frequency; respectively, Broder).

Regarding Claims 17, and 53, the combination of Broder in view of Moffat discloses a method, wherein wherein the boundary posting includes a document identifier having the lower value in the range of values and the block of postings follow the first skip entry in the inverted list (Fig. 27, item 5, Page 16 and 17, [0301], lines 10 – 18, Broder[10]; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

---

there is DID < id (the number of postings exceeds the number of documents entered), then the iterator returns the special posting mentioned above.

[10] According to Broder, the pivot DID is the smallest DID that might be a candidate (Page 17, [0301], lines 17 – 18). In addition, Broder discloses that this pivot term goes through an "if" statement which finds a first pivot term with UB (upper bound) greater than the threshold (Fig. 27, item 5). This implies that the

Regarding Claims 18, and 54, the combination of Broder in view of Moffat discloses a method, wherein the first skip entry further includes information to locate the next skip entry in the inverted list (Fig. 27, items 13 and 22, Broder[11] and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

Regarding Claims 19, and 55, the combination of Broder in view of Moffat discloses a method, wherein the boundary posting includes a document identifier having the higher value in the range of values and the block of postings precede the first skip entry in the inverted list (Page 17, [0302], lines 22 – 26, Broder[12]; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

"next" method will return the smallest possible document number following the last one. In addition, the smallest document number corresponds to the lower value claimed.

[11] Posting[aterm] (located in the function next()) represents the information to locate next skip.

[12] The next() function iterates trough the list and selects from the preceding terms the term with the location greater (largest document number of documents) than the pivot location. Wherein the location

Regarding Claims 20, and 56, the combination of Broder in view of Moffat discloses a method, wherein the first skip entry further includes information to locate the next skip entry in the inverted list (Fig. 27, items 13 and 22, Broder[13]; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

Regarding Claims 21, and 57, the combination of Broder in view of Moffat discloses a method, wherein substantially all of the index terms occur in x documents or fewer (Page 15, [0257], lines 14 – 16, Broder).

Regarding Claims 22, and 58, the combination of Broder in view of Moffat discloses a method, wherein approximately 80 to 90% of the index terms occur in x documents or fewer (Page 17, [0307], lines 1 – 3, top n results, Broder).

Regarding Claims 23, and 59, the combination of Broder in view of Moffat discloses a method, wherein for each inverted list, if the number of postings exceeds x, further including a skip entry before the first posting in the inverted list (Page 15, [0257],

---

greater (largest document number of documents) than the pivot location corresponds to the higher value

lines 12 – 20, the result is inserted, Broder; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

Regarding Claims 24, and 60, the combination of Broder in view of Moffat discloses a method, wherein for each inverted list, if the number of postings exceeds x (Page 14, [0245], lines 7 – 10, Broder[14]), further including a skip entry after the last posting in the inverted list (Page 14, [0245], lines 10 – 12, Broder; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

Regarding Claim 25, the combination of Broder in view of Moffat discloses an inverted index for a collection of documents (Page 9, [0162], lines 1 – 3, Broder), each

---

as claimed.
[13] Posting[aterm] (located in the function next()) represents the information to locate next skip.
[14]Broder discloses that the method next(id) (page 14, [0245], lines 8 – 10). This method states that if there is **not** such document which DID >=id (where DID = number of documents, and id = number of posting), then the term iterator returns a special posting that is larger than all the existing DIDs. This implies that, if there is DID < id (the number of postings exceeds the number of documents entered), then the iterator returns the special posting mentioned above.

document comprising one or more index terms (Page 14, [0244], lines 3 – 5, Broder), the inverted index comprising:

an inverted list for each index term in the collection of documents (Page 14, [0244], lines 2 – 3, Broder); and

one or more inverted lists including a quantity of postings (Page 14, [0244], lines 2 – 5, Broder) that exceeds a value x (Page 14, [0245], lines 7 – 10, Broder), a skip entry inserted after the xth posting (Page 14, [0245], lines 10 – 12, Broder and also see Page 18, Section 4.4 Implementation, 1st paragraph, "…**inverted files were built with skips inserted** into each list assuming that k had some fixed value …", Moffat), and one or more additional skip entries inserted thereafter at intervals of every yth posting (Page 15, [0277], lines 1 – 6, Broder[15]; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat and also see Page 14, 4th paragraph, "a Golomb code is used for both the **inverted lists and the two components of the skips inserted into each**.", Moffat), where the value x is determined as an integer corresponding to a quantity of documents including at least a majority of the index terms (Page 17 and 18, [0307] and [0314], lines 1 – 3 and 3 – 9;

[15] The cursor is advanced to the position of k value. There is a skip at the kth value.

respectively, Broder[16]) and the value y is determined as an integer and does not exceed the value x (Page 15, [0277], lines 1 – 2, Broder[17]);

wherein:

a posting includes a document identifier identifying a document in the collection of documents (Page 14, [0244], lines 1 – 6, document unique identifier DID, Broder);

a skip entry includes a document identifier that is included in a boundary posting of a block of postings immediately adjacent to the skip entry in the inverted list (Page 15, [0277], lines 1 – 6, Border[18]; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat), where a block of postings includes postings having document identifiers ranging from a lower to an upper value and where a boundary posting is a posting having a document identifier of either the lower or the upper value (Page 16 and 18, [0286] and [0314], lines 5 – 9 and 3 – 9; respectively, Broder).

The combination of Broder in view of Moffat also discloses:

---

[16] Wherein the top "n" documents corresponds to the documents with the majority of index terms claimed; specifically "n" corresponds to the value x claimed. The scoring procedure utilized by Broder (as disclosed in detailed in Page 16, [0286], lines 5 – 6, Broder) including terms associations with upper bounds on its maximal contribution to documents scores; wherein the maximal contributing of index terms corresponds to the majority of the index terms as claimed. Additionally, the top n scoring documents corresponds to a quantity of documents as claimed.

[17] Y would be the **smallest index k**. And x would be the top n documents (the size of the heap) (Page 15, [0257], lines 15 – 16).

wherein y does not equal x, such that the first and second skip entries are provided at different intervals (Fig. 13, see jump/skip entry after "software" from 5-12, and jump/skip entry after "is" from 14 - 15; as shown in Fig 13, the intervals are different; Broder; and also Page 20, lines 9 – 15; "Another way to include skipping in an index is to vary the L parameter for each inverted list rather than use the same value…"; Moffat).

Regarding Claim 26, the combination of Broder in view of Moffat discloses an inverted index, wherein each posting further includes position identifier identifying a position of the index term in the document (Page 14, [0244], lines 8 – 9, offsets of occurrences, Broder), and a frequency of the index term occurring in the document (Page 14 and 18, [0244] and [0316], lines 7 – 8 and 9 – 11, number of occurrences of the terms/ frequency; respectively, Broder).

Regarding Claim 27, the combination of Broder in view of Moffat discloses an inverted index, wherein the boundary posting includes a document identifier having the lower value in the range of values and the block of postings follow the first skip entry in the inverted list (Fig. 27, item 5, Page 16 and 17, [0301], lines 10 – 18, Broder[19]; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization

---

[18] Wherein doc_i corresponds to the document identifier claimed.
[19] According to Broder, the pivot DID is the smallest DID that might be a candidate (Page 17, [0301], lines 17 – 18). In addition, Broder discloses that this pivot term goes through an "if" statement which finds a first pivot term with UB (upper bound) greater than the threshold (Fig. 27, item 5). This implies that the "next" method will return the smallest possible document number following the last one. In addition, the smallest document number corresponds to the lower value claimed.

points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

Regarding Claim 28, the combination of Broder in view of Moffat discloses an inverted index, wherein the first skip entry further includes information to locate the next skip entry in the inverted list (Fig. 27, items 13 and 22, Broder[20]; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

Regarding Claim 29, the combination of Broder in view of Moffat discloses an inverted index, wherein the boundary posting includes a document identifier having the higher value in the range of values and the block of postings precede the first skip entry in the inverted list (Page 17, [0302], lines 22 – 26, Broder[21]; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…",

[20] Posting[aterm] (located in the function next()) represents the information to locate next skip.
[21] The next() function iterates trough the list and selects from the preceding terms the term with the location greater (largest document number of documents) than the pivot location. Wherein the location

"…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

Regarding Claim 30, the combination of Broder in view of Moffat discloses an inverted index, as disclosed above, wherein the first skip entry further includes information to locate the next skip entry in the inverted list (Fig. 27, items 13 and 22, Broder[22]; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

Regarding Claim 32, the combination of Broder in view of Moffat discloses an inverted index, wherein substantially all of the index terms occur in x documents or fewer (Page 15, [0257], lines 14 – 16, Broder).

Regarding Claim 33, the combination of Broder in view of Moffat discloses an inverted index, wherein approximately 80% of the index terms occur in x documents (Page 17, [0307], lines 1 – 3, top n results, Broder).

---

greater (largest document number of documents) than the pivot location corresponds to the higher value as claimed.

[22] Posting[aterm] (located in the function next()) represents the information to locate next skip.

Regarding Claim 34, the combination of Broder in view of Moffat discloses an inverted index, wherein the collection of one or more documents includes one or more binary files, data tables, source code files, text documents or combinations thereof (Page 9, [0158], lines 1 – 13, Broder).

Regarding Claim 35, the combination of Broder in view of Moffat discloses an inverted index, wherein the one or more inverted lists further include a skip entry before the first posting in the inverted list (Page 15, [0257], lines 12 – 20, the result is inserted, Broder; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

Regarding Claim 36, the combination of Broder in view of Moffat discloses an inverted index, wherein the one or more inverted lists further include a skip entry after the last posting in the inverted list (Page 14, [0245], lines 10 – 12, Broder; and Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, "if synchronization points-additional locations at which decoding can commence – are introduced into the compressed inverted list…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat).

Regarding Claim 8, and 44, the combination of Broder in view of Moffat discloses all the limitations as discussed above including: wherein x is selected and y is selected, wherein y is not selected to have the same value as x (Page 15, [0277], lines 1 – 2, k be the smallest index, Broder; and Page 13, Section 4.1 Skipping, paragraph 1, "…suppose that p1 synchronisation points are allowed. The index into the inverted list contains p1 "document number, bit address" pairs, and can itself be stored as a compressed sequence of "difference in document number, difference in bit address run lengths…", Moffat) and selecting ranges (Page 7, [0140], "discovering a range…"; Broder). However, Broder does not expressly disclose the specific ranges of 256 to 512 and of 128 to 256. It would have been an obvious matter of design choice to include the ranges of 256 to 512 and 128 to 256, since applicant has not disclosed that ranges of 256 to 512 and 128 to 256 solves any stated problem or is for any particular purpose and it appears that the invention would perform equally well with ranges of 256 to 512 and 128 to 256.

Regarding Claim 15, and 51, the combination of Broder in view of Moffat discloses an article comprising a machine-readable medium storing instructions operable to cause one or more machines to perform operations comprising:

receiving a collection of documents, each document comprising one or more index terms (Page 14, [0244], lines 3 – 5, Broder);

determining a first value x representative of first location for inserting a first skip entry in an inverted index (Page 14, Section 4.1: Skipping, 1st – 3rd paragraph in that page, ""if synchronization points-additional locations at which decoding can commence – **are introduced into the compressed inverted list.** For example, suppose that p1 synchronisation points are allowed. Then the index into the inverted list contains p1 'document number, bit address' pairs, and can itself be stored as a compressed sequence of 'difference in document number, difference in bit address'…", "…To access the compressed list to see if document *d* appears, the first skip is decoded to obtain the address a2 of the second skip, which is also decoded…"; Moffat), wherein at least a majority of the index terms occur in x documents (Page 17 and 18, [0307] and [0314], lines 1 – 3 and 3 – 9; respectively, Broder[23]) and x is an integer (Page 10, [0182], lines 6 – 8, Broder);

determining a second value y representative of second location for inserting a second skip entry in an inverted index (Page 14, 1st and 2nd paragraph, "…the inverted list becomes a sequence of blocks of three pairs each, with skips separating the blocks…", and "…a2 is the address of the first bit of the second skip pair, a3 is the address of the first bit of the third skip, and so on…", Moffat and also see Page 14, 4th paragraph, "a Golomb code is used for both the **inverted lists and the two components of the skips inserted into each**.", Moffat), wherein **y** does not exceed

[23] Wherein the top "n" documents corresponds to the documents with the majority of index terms claimed; specifically "n" corresponds to the value x claimed. The scoring procedure utilized by Broder (as disclosed in detailed in Page 16, [0286], lines 5 – 6, Broder) including terms associations with upper bounds on its maximal contribution to documents scores; wherein the maximal contributing of index terms corresponds to the majority of the index terms as claimed.

the value of x (Page 15, [0277], lines 1 – 2, **k be the smallest** index, Broder) and is an integer (Page 10, [0182], lines 6 – 8, Broder);

generating the inverted index for the collection of documents (Page 9, [0162], lines 1 – 3, Broder; and Page 1, Abstract, "…an internal index in each inverted list…", Moffat), the inverted index including an inverted list for each of the index terms (Page 14, [0244], lines 2 – 3, Broder), each inverted list including at least one posting (Page 14, [0244], lines 2 – 5, Broder) and, if the number of postings exceeds x (Page 14, [0245], lines 7 – 10, Broder[24]), further including the first skip entry inserted after the xth posting (Page 14, [0245], lines 10 – 12, Broder; and Page 13, Section: 4.1 Skipping, "…When k<<p, faster performance is possible if synchronization points-additional locations at which decoding can commence-are introduced into the compressed inverted list…", Moffat and also see Page 18, Section 4.4 Implementation, 1st paragraph, "…**inverted files were built with skips inserted** into each list assuming that k had some fixed value …", Moffat) and one or more second skip entries inserted thereafter at intervals of every yth posting (Page 15, [0277], lines 1 – 6, Broder[25]; and Page 14, 1st and 2nd paragraph, "…the inverted list becomes a sequence of blocks of three pairs each, with skips separating the blocks…", and "…a2 is the address of the first bit of the second skip pair, a3 is the address of the first bit of the third skip, and so

[24]Broder discloses that the method next(id) (page 14, [0245], lines 8 – 10). This method states that if there is **not** such document which DID >=id (where DID = number of documents, and id = number of posting), then the term iterator returns a special posting that is larger than all the existing DIDs. This implies that, if there is DID < id (the number of postings exceeds the number of documents entered), then the iterator returns the special posting mentioned above.

[25] The cursor is advanced to the position of k value. There is a skip entry at the kth value.

on…", Moffat and also see Page 14, 4th paragraph, "a Golomb code is used for both the **inverted lists and the two components of the skips inserted into each**.", Moffat).

wherein:

the at least one posting includes a document identifier identifying a document in the collection of documents (Page 14, [0244], lines 1 – 6, document unique identifier DID, Broder; and Page 3, Section 2: Document Database, 1st paragraph, "Each document is known by a unique indetifier…", Moffat);

the first and second skip entries include a document identifier that is included in a boundary posting of a block of postings immediately adjacent to the skip entry in the inverted list (Page 15, [0277], lines 1 – 6, Border[26]; and Page 13, Section 4.1 Skipping, 6th paragraph, "… For example, suppose that p1 synchronisation points are allowed. Then the index into the inverted list contains p1 'document number, bit address' pairs, and can itself be stored as a compressed sequence of 'difference in document number, difference in bit address'…"; Moffat), where a block of postings includes postings having document identifiers ranging from a lower to an upper value and where a boundary posting is a posting having a document identifier of either the lower or the upper value (Page 16 and 18, [0286] and [0314], lines 5 – 9 and 3 – 9; respectively, Broder).

The combination of Broder in view of Moffat also discloses:

wherein y does not equal x, such that the first and second skip entries are provided at different intervals (Fig. 13, see jump/skip entry after "software" from 5-12,

[26] Wherein doc_i corresponds to the document identifier claimed.

and jump/skip entry after "is" from 14 - 15; as shown in Fig 13, the intervals are different; Broder; and also Page 20, lines 9 – 15; "Another way to include skipping in an index is to vary the L parameter for each inverted list rather than use the same value…"; Moffat).

The combination of Broder in view of Moffat discloses all the limitations as discussed above including: selecting ranges (Page 7, [0140], "discovering a range…"; Broder). However, Broder does not expressly disclose the specific ranges of 256 to 512 and of 128 to 256. It would have been an obvious matter of design choice to include the ranges of 256 to 512 and 128 to 256, since applicant has not disclosed that ranges of 256 to 512 and 128 to 256 solves any stated problem or is for any particular purpose and it appears that the invention would perform equally well with ranges of 256 to 512 and 128 to 256.

Regarding Claim 31, the combination of Broder in view of Moffat discloses an inverted index, the combination of Broder in view of Moffat discloses all the limitations as discussed above including: wherein x is selected and y is selected, wherein y is cannot selected to have the same value as x (Page 15, [0277], lines 1 – 2, k be the smallest index, Broder; and Page 13, Section 4.1 Skipping, paragraph 1, "…suppose that p1 synchronisation points are allowed. The index into the inverted list contains p1 "document number, bit address" pairs, and can itself be stored as a compressed sequence of "difference in document number, difference in bit address run lengths…", Moffat) and selecting ranges (Page 7, [0140], "discovering a range…"; Broder).

However, Broder does not expressly disclose the specific ranges of 256 to 512 and of 128 to 256. It would have been an obvious matter of design choice to include the ranges of 256 to 512 and 128 to 256, since applicant has not disclosed that ranges of 256 to 512 and 128 to 256 solves any stated problem or is for any particular purpose and it appears that the invention would perform equally well with ranges of 256 to 512 and 128 to 256.

Prior art Made of Record

1. Broder et al. (US Patent Application Pub. No. 2004/0243560 A1) discloses a system, method and computer program product for performing unstructured information management and automatic text analysis, including an annotation inverted file system facilitating indexing and searching.

2. Antoshenkov (US Patent No. 6,439,783 B1) discloses a range-based query optimizer.

3. Huynh et al. (US Patent No. 5,539,899) discloses a system and method for handling a segmented program in a memory for a multitasking data processing system utilizing paged virtual storage.

4. Young et al. (US Patent No. 5,838,950) discloses a method of operation of a host adapter integrated circuit.

5. Moffat et al. (Moffat hereinafter) (Non-Patent Literature: "Self-Indexing Inverted Files for Fast Text Retrieval"; Alistar Moffat, and Justin Zobel; February 1994, 1994 Australian Database Conference, and 1994 IEEE Conference on Data Engineering)

Points Of Contact

Any inquiry concerning this communication or earlier communications from the examiner should be directed to GIOVANNA COLAN whose telephone number is (571)272-2752. The examiner can normally be reached on 8:30 am - 5:00 pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, John Breene can be reached on (571) 272-4107. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

Giovanna Colan
Examiner
Art Unit 2162
March 24, 2008

/Jean M Corrielus/
Primary Examiner, Art Unit 2162